

10266)

10266)

RESULT 8
 ATHRIENUCP
 LOCUS 1160 bp mRNA linear PLN 27-APR-1993
 DEFINITION Arabidopsis thaliana ribonucleoprotein mRNA, complete cds.
 ACCESSION M98340
 VERSION M98340.1 GI:166843
 KEYWORDS homologue; ribonucleoprotein; splicing factor; splicing factor SF-2.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1160)
 Lazar G.
 Unpublished (1992)
 Original source text: Arabidopsis thaliana (strain Landsberg erecta) (library: lambda gt11) root cDNA to mRNA.
 Location/Qualifiers
 1..1160
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Landsberg erecta"
 FEATURES
 source

db 691 AGCGTGAGCCGAGCAGGAGGAGGAGGAGC 723
RESULT 9

/db_xref="taxon:3702"
/tissue_type="root"
/tissue_lib="Lambda gt11"
31..942
/note="human splicing factor SF-2 homologue"
/codon_start=1
/product="ribonucleoprotein"
/protein_id="AAA32856.1"
/db_xref="GI:166844"
/translation="MSSRSRTVTVGNLPGDINREVEDELPSKYGVVQIDLVPPRP
PGYAFVDFDARDADAIHGRDGYDFDEHRLRYELAHGRRSSDDTRGSFNGGRRGG
RGDGGSRFPVSRSEFRVLTWLASASQDLKDHIAKGDDVCFQVYVDRAGTTGV
VDYTCYDMKYALKLDDTFRNAPSNVYRVREYDNRKDSRSPGRSRSYKSRSR
GRSVRSRSRSPKAKSPKAKSTSRSPGRSRSRSPRSPRSPRSPRSPRSPRSP
SVQEGSKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPKSPK
BASE COUNT 300 a 233 c 318 g 309 t
ORIGIN

Alignment Scores:
Pred. No.: 2.34e-15 Length: 1160
Score: 139.00 Matches: 144
Percent Similarity: 66.67% Conservative: 10
Best Local Similarity: 62.34% Mismatches: 67
Query Match: 62.61% Indels: 10
DB: 8 Gaps: 1

US-10-014-927-19mod_copy_1_222 (1-222) x ATHRIENUCP (1-1160)
QY 1 MetSerSerArgTTPAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20
Db 31 ATGAGCAGTCGTTTCGAGTAGAACCGTGATCGCGAAACCTTCCTGGCGATATCCGTGAG 90
QY 21 CysGluValGluAspLeuPheTyrIleTyrGlyProIleValAspIleAspLeuLysIle 40
Db 91 AGAGAGGTGCAAGATTGTTTCAAGTATGAGACTGTGTTCAAAATTGATTGTAAGGTT 150
QY 41 ProProArgProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp 60
Db 151 CCTCCAGGCTCTCTGTTATGATTCATTCCTGATTTGATGATGCTCGGATGCTGAAGAT 210
QY 61 AlalleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla 80
Db 211 GCTATTTCATGTCGTGATGGTATGATTCATGATGGCATCGTTTGAAGGTTGAATTGGCG 270
QY 81 HisGlyGlyArgArg-----***** 90
Db 271 CATGGTGGAGCGGTTTCATCATGATGATCTCGGGTAGTTTCAATGTTGGTGGCGGTGGT 330
QY 91 *****SerTyrSerAlaSerArgAlaProSerArgArgSerAspTyrArg 110
Db 331 GGTGGTGGCGCGCGTGTGATGCTGATGCTGATGCTGCGCCATCTAGGAGATCAGAGTTTCGT 390
QY 111 ValLeuValThrGlyLeuProProSerAlaSerThrGlnAspLeuLysAspHisMetArg 130
Db 391 GTTCTAGTCACATGGCTTCT 450
QY 131 LysAlaGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGlyVal 150
Db 451 AAAGAGGGGAGTCTGTTTCTGCGAGTGTACCGTGTGCTAGAGGAGCAACTGGAGTT 510
QY 151 ValAspTyrSerAsnTyrAspAspMetLysTyrAlaIleArgLysLeuAspAlaThrGlu 170
Db 511 GTTGATTACACCTGCTATCAGGACATGAAGTATGCGCTGAAAGAGCTCGACACACAGAG 570
QY 171 PheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgSerVal 190
Db 571 TTTTCGAAATCGCTTTTCGAAATGGAATATGTCGGGTTAGAGATATGATTCAAGGAGGAT 630
QY 191 SerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgGlyProSer 210
Db 631 TCTAGAGTCTAGCGGGAGAGATCCTATTCTAGAGCGCGCGAGCGCGCGCGCGCGCGGAGGA 690
QY 211 CysSerTyrSerSerLysSerArgSerValSer 221

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model
Run on: February 4, 2004, 21:21:59 ; Search time 3510 Seconds
(without alignment)
2587.446 Million cell updates/sec

Title: US-10-014-927-19MOD_COPY_1_222
Perfect score: 222
Sequence: 1 MSSRWNRITVGNLFGDIRK.....RSRRGPGSCSYSSKSRVSVP 222

Scoring table: UNITARY2
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/BAUM927/runat_04022004_131333_2706/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LCOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=unitary2 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pco -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=BAUM927 @CGN 1.1 3508 @runat_04022004_131333_2706 -NCPU=6 -ICPU=3
-NO_XMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DISPATCH=100 -LONLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_hg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_ov.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rdd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	222	100.0	838	8	AY150486 Arabidops
2	207	93.2	762	6	AX506504 Sequence
3	160	72.1	933	8	AY050912 Arabidops
4	143	64.4	858	8	BT006316 Arabidops
5	143	64.4	1229	8	AK118379 Arabidops
6	143	64.4	1232	8	AY085920 Arabidops
7	143	64.4	1540	8	AY128338 Arabidops
8	139	62.6	1160	8	M98340 Arabidops
9	131.5	59.2	1523	8	AY056185 Arabidops
10	131.5	59.2	2059	8	AK118074 Arabidops
11	95	42.8	1428	6	BD094070 Shear str
12	95	42.8	1428	9	M69040 Human SF2p3
13	95	42.8	2708	9	BC010264 Homo sapi
14	95	42.8	2765	9	AB062124 Homo sapi
15	95	42.8	2878	10	BC046773 Mus muscu
16	89	40.1	1618	5	BC042354 Xenopus l
17	88	39.6	1099	3	AK115991 Ciona int
18	87.5	39.4	1069	6	AX410696 Sequence
19	87.5	39.4	1069	9	HSU30825 Human spli
20	87.5	39.4	1193	10	BC012217 Mus muscu
21	87.5	39.4	1584	6	AX305513 Sequence
22	87.5	39.4	1584	10	X66091 M.musculus
23	87.5	39.4	1717	9	M72709 Human alter
24	86.5	39.0	2369	5	BC046679 Xenopus l
25	73.5	33.1	1391	3	AF232773 Drosophil
26	73.5	33.1	1395	3	AY051919 Drosophil
27	68	30.6	809	3	AF242767 Caenorhab
28	68	30.6	160415	2	AC140961 Papio anu
29	65.5	29.5	343050	3	PFA929353 Plasmodiu
30	63	28.4	495	6	AX306308 Sequence
31	63	28.4	497	6	BD032870 Sequence
32	62	27.9	1485	9	BT007415 Homo sapi
33	62	27.9	1485	12	BT008254 Synthetic
34	62	27.9	2076	9	HUNSRP75A Homo pre-m
35	62	27.9	2167	9	BC002781 Homo sapi
36	62	27.9	2204	10	BC019437 Mus muscu
37	59	26.6	1845	5	BC046895 Danio rer
38	58	26.1	3366	5	BC046668 Xenopus l
39	58	26.1	5164	8	AT131214 Arabidops
40	58	26.1	59261	8	T12M4 Arabidops
41	57.5	25.9	3081	5	BC044265 Xenopus l
42	57.5	25.9	3130	10	BC026944 Mus muscu
43	57	25.7	1579	10	RATRHS Rattus norv
44	57	25.7	181545	10	AL626774 Mouse DNA
45	56	25.2	4044	6	AX040661 Sequence

ALIGNMENTS

RESULT 1

```

AY150486      838 bp      mRNA      linear      PLN 23-SEP-2002
LOCUS      Arabidopsis thaliana putative SP2/ASF splicing modulator Srp30
DEFINITION      (Atg09140) mRNA, complete cds.

ACCESSION      AY150486
VERSION      AY150486.1 GI:23297698
KEYWORDS      FLI_CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 838)
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J.,
Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T.,
Davis,R.W., Ecker,J.R. and Theologis,A.
Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
2 (bases 1 to 838)
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.K., Wong,C.,
Wu,H.C., Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J.,
Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T.,
Davis,R.W., Ecker,J.R. and Theologis,A.
Direct Submission
Submitted (13-SEP-2002) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.
FEATURES             source
     Location/Qualifiers
         1..838
            /organism="Arabidopsis thaliana"
            /mol_type="mRNA"
            /db_xref="taxon:3702"
            /chromosome="1"
            /clone="C10S028"
            /note="This clone is in pUNI 51.
            ecotype: Columbia"
         1..838
            /gene="At1g09140"
         1..807
            /gene="At1g09140"
         /codon_start=1
         /evidence=experimental
         /product="putative SP2/ASF splicing modulator Srp30"
         /protein_id="AA13011.1"
         /db_xref="GI:23297699"
         /translation="MSRRNRRTIVGNLPDRLKCEVEDLPYKGPVIDIDLKIPPRP
PGYAFVEFPRDADDAIYGRDGYDFDGLRVEIAHGGRRFSPVDYSSYSASRA
PSRRDYRLVLTGLPPSASQDLKHKRAGDVCFSEVDFDKMGSGVVDYNDMK
YAIRKLDATFRNAPFSAYTRVREYSRVSRSPPDCKSRSRSGSPGSCYSKSR
SVSPARSIPRSRPLSRSLYSVRSRSQSRKSRSRSRSPVSPVSG"
         808..838
            /gene="At1g09140"

3'UTR
BASE COUNT      212 a      176 c      219 g      231 t
ORIGIN
Alignment Scores:
Pred No.:      3,69e-32      Length:      838
Score:      222.00      Matches:      212
Percent Similarity:      100.00%      Conservative:      10
Best local Similarity:      95.50%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      8      Gaps:      0

US-10-014-927-19MOD_COPY_1_222 (1-222) x AY150486 (1-838)

QY      1 MetSerSerArgTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20
DB      1 ATGAGTAGCCGATGGATCGTACGATCTAGCTTGGGAATTGGCTGGAGATATCGCAAG 60

```

```

QY      21 CysGluValGluAspLeuPheTyrIleTyrGlyProIleValAspIleAspLeuIle 40
DB      61 TGTGAGGTGAAGATCTCTTCTACAGTATGACCAATTGTGACATTGATTGAAGATT 120

QY      41 ProProArgProProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAsp 60
DB      121 CCACCGAGACCTCTCGTTATGCTTTCGAGTTGAAGATCCTCGTGATGCACGAC 180

QY      61 AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla 80
DB      181 GCATTTATGGAGTGTATGTTATGTTTGGTTCGACTTCGACTTCGGGTTCGAGATTGCA 240

QY      81 HisGlyGlyArgArg*****SerTyrSerAlaSer 100
DB      241 CATGGTGGTGGTGAATTTTTCACCATCAGTTGATAGGTACAGCAGCAGCTACAGTGC 300

QY      101 ArgAlaProSerArgArgSerArgTyrArgValLeuValThrGlyLeuProProSerAla 120
DB      301 CGTGCACCTTCAAGACGCTCTGACTACCGCGGTGTGACCGGATTCACCGCTTCTCT 360

QY      121 SerTrpGlnAspLeuLysAspHisMetArgLysAlaGlyAspValCysPheSerGluVal 140
DB      361 TCGTGGCAGGACCTTAAGGATCACATGCGCAAGCTGGAGATGTCTGCTTCTCTGAAGTT 420

QY      141 PheProAspArgLysGlyMetSerGlyValValAspTyrSerAsnTyrAspAspMetLys 160
DB      421 TTCCTCGACCGTGAAGCATGTCTGGGTGTGGATTATAGCAACTATGATGATATGAAG 480

QY      161 TyrAlaIleArgLysLeuAspAlaThrGluPheArgAsnAlaPheSerSerAlaTyrIle 180
DB      481 TAGCAATPAGGAATTTGATGCCACTGANTTCGAATTCCTTCTCTAGTCTTATATA 540

QY      181 ArgValArgGluTyrGluSerArgSerValSerArgSerProAspAspSerIleSerTyr 200
DB      541 CGGGTCAGGGAATATGAGTCGAGGAGTGTGAGTCAAGCCAGATGATTTCTAAAGCTAT 600

QY      201 ArgSerArgSerArgSerArgGlyProSerCysSerTyrSerSerIleSerArgSerVal 220
DB      601 AGAAGCAGGAGTCGGAGCCGTGTCAGCTGTAGCTATAGTAGCAGAGGAGGTGTG 660

QY      221 SerPro 222
DB      661 TCACCT 666

RESULT 2
AX506504      762 bp      DNA      linear      PAT 27-SEP-2002
LOCUS      Sequence 1199 from Patent WO0216655.
DEFINITION      AX506504
ACCESSION      AX506504
VERSION      AX506504.1 GI:23387741
KEYWORDS      Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1
Harper,J.F., Kreps,J., Wang,X. and Zhu,T.
Stress-regulated genes of plants, transgenic plants containing
same, and methods of use
Patent: WO 0216655-A 1199 28-FEB-2002;
The Scripps Research Institute (US); Syngenta Participations AG
(CH)
Location/Qualifiers
     1..762
        /organism="Arabidopsis thaliana"
        /mol_type="genomic DNA"
        /db_xref="taxon:3702"
BASE COUNT      189 a      161 c      197 g      215 t
ORIGIN
Alignment Scores:
Pred. No.:      2.86e-29      Length:      762

```

Score: 207.00 Matches: 197
Percent Similarity: 100.00% Conservative: 10
Best Local Similarity: 95.17% Mismatches: 0
Query Match: 93.24% Indels: 0
DB: 6 Gaps: 0

US-10-014-927-19MOD_COPY_1_222 (1-222) x AX506504 (1-762)

QY	1	MetSerSerArgTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys	20
DB	1	ATCAGTAGCCGATGGATCGATCGATCTACGTTGGAAATTGCTGGAGATTCGCAAG	60
QY	21	CysGluValGluAspLeuPheTyrLysTyrGlyProIleValAspIleAspLysIle	40
DB	61	TGTGAGGTTGAAGATCTCTTCAACAAGTATGGACCAATTGTGACATTTGAAGATT	120
QY	41	ProProArgProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAsp	60
DB	121	CCACCGAGACCTCTGTTGATGCTTTGCGAGTTTGAAGATCCTCGTATGACAGCAT	180
QY	61	AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla	80
DB	181	GCAATTTATGAGTGAATGTTATGATTTGATGGTGTGCGATTCGGGTTGAGATTGCA	240
QY	81	HisGlyGlyArgArg*****SerTyrSerAlaSer	100
DB	241	CATGTTGTTGATGATTTTCCACCATCAGTTGATAGGTACAGCAGCAGCTACAGTGCAGC	300
QY	101	ArgAlaProSerArgArgSerArgTyrArgValLeuValThrGlyLeuProProSerAla	120
DB	301	CGTGACCTTCAAGACGCTCTGACTACCGGCTGCTTGTGACCGGATTCGCGCTTCGT	360
QY	121	SerTrpGlnAspLeuLysAspHisMetArgLysAlaGlyAspValCysPheSerGluVal	140
DB	361	TCGTGGCAGACCTTAAAGATCATCGCGAAGCTGGAGATGCTCTCTCTCTGAGTT	420
QY	141	PheProAspArgLysGlyMetSerGlyValValAspTyrSerAsnTyrAspMetLys	160
DB	421	TTCCCTGACGTAAGACATGCTCGGGTTGTGATATAGCAACATGATGATGATGAAG	480
QY	161	TyrAlaIleArgLysLeuAspAlaThrGluPheArgAsnAlaPheSerSerAlaTyrIle	180
DB	481	TACGCAATAGGAACATGATGACCTGCACTGCAATTTGCAAAATGCTTCTCTAGTGTATATA	540
QY	181	ArgValArgGluTyrGluSerArgSerValSerArgSerProAspSerLysSerTyr	200
DB	541	CGGCTGAGGAATATGATGATGAGGAGTGTGATGCGAGCCGAGATTCATTAAGCTAT	600
QY	201	ArgSerArgSerArgSerArg	207
DB	601	AGAAAGCAGGATCGGAGCCGT	621

RESULT 3
AY050912 933 bp mRNA linear PLN 18-SEP-2002
LOCUS Arabidopsis thaliana putative SP2/ASF splicing modulator Srp30
DEFINITION protein (At1g09140) mRNA, partial cds.
ACCESSION AY050912
VERSION AY050912.1 GI:15292956
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS
Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K., Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Arabidopsis Full Length cDNA Clones

Unpublished
2 (bases 1 to 933)
Yamada, K., Liu, S.X., Pham, P.K., Banh, J., Banno, F., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Chodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission
Submitted (24-JUL-2001) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PSEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada, K., Liu, S.X., Pham, P.K., Banh, J., Banno, F., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S., Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W., Ecker, J.R. and Theologis, A.
Yamada, K. (SP/PSEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PSEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

Location/Qualifiers
1..933
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="1"
/clone="RAFL08-12-105 (R11038)"
/note="This clone is in a modified pBluescript vector (FLC-1) as a BamHI/XhoI insert."
ecotype: Columbia
c1..933
/gene="At1g09140"
c1..626
/gene="At1g09140"
/codon_start=3
/evidence=experimental
/product="putative SP2/ASF splicing modulator Srp30 protein"
/protein_id="AAK33589.2"
/db_xref="GI:21954085"
/translation="TGGSDGYDFDGCRLRVEIAHGRRFSPVDYSSYSASRAPSRSDHVLVTGLPPASMQDLADHMKAGDVCSEVFPDRKMGSGVVDYNDYDKYAIKLDATFRNAPSSAIKRVREYSESRSPDDSKSYSESRSPGSCSYSKSRSVSPARSIPSRPLSRSLYSVSRSGRSKSRTRSRNSPVSPIVG"
misc_difference 1
/gene="At1g09140"
/note="compared to genomic sequence"
/replace="c"
misc_difference 148
/gene="At1g09140"
/note="compared to genomic sequence"
/replace="g"
misc_difference 579
/gene="At1g09140"
/note="compared to genomic sequence resulting in an amino acid sequence difference"
/replace="t"

Db 61 AGAGAGGTCGAGATTGTTTCAGTAGTATGACCTGTTGTTTCACATTGATTGAAGGTT 120

Qy 41 ProProArgProProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp 60

Db 121 CTTCCAAAGGCTCTGCTTATGATCGTTGATGTTGATGATGCTCGGATCTCAAGAT 180

Qy 61 AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuValGluIleAla 80

Db 181 GCTATTTCATGCTGCTGATGCTATGATGCTGATGCTGATGCTGATGCTGATGCTG 240

Qy 81 HisGlyGlyArgArg----- 90

Db 241 CATGTTGGAGGCTTTCATCAGTATGATCTCGGGGTAGTTTCAATGGTGGTGGTGGT 300

Qy 91 *****SerTyrSerAlaSerAlaProSerArgAlaProSerArgSerAspTyrArg 110

Db 301 GGTGGTGGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360

Qy 111 ValLeuValThrGlyLeuProProSerAlaSerTyrGlnAspLeuIleAspHisMetArg 130

Db 361 GTTCTAGTCACAGGCTTGGCTTCTCATCTGCTTCTTGGCAAGATCTCAAGATCACATGCGT 420

Qy 131 LysAlaGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGlyVal 150

Db 421 AAAGGAGGCGATCTCTGTTCTCGCAAGTGTACCGTGTGCTAGAGGGGCAACTGGAGTT 480

Qy 151 ValAspTyrSerAsnTyrAspMetLysTyrAlaIleArgLysLeuAspAlaThrGlu 170

Db 481 GTTGATTACCTGCTATGAGGACATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

Qy 171 PheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgSerVal 190

Db 541 TTTTGAATGCGTTTTCGAATGATGATGCTCGGGTTAGAGATATGATTCAGGAAGGAT 600

Qy 191 SerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgSerArg 210

Db 601 TCTAGGAGTCTTACCGCGGAGATCTTCTTAAAGCGCGAGCGCGCGCGCGCGCGCG 660

Qy 211 CysSerTyrSerSerLysSerArgSerValSer 221

Db 661 AGCGTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 693

RESULT 5

AK118379

LOCUS

DEFINITION

Arabidopsis thaliana At1g02840 mRNA for putative ribonucleoprotein SF-2, complete cds, clone: RAF19-64-H16.

ACCESSION

AK118379

VERSION

AK118379.1 GI:26451793

KEYWORDS

Full cDNA, CAP trapper

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

2 Arabidopsis thaliana full-length cDNA

Published Only in Database (2002)

2 (bases 1 to 1229)

3 Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

Direct Submission

Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: msek@gs.riken.go.jp).

URL: <http://pfweb.gsc.riken.go.jp>, Tel: 81-45-503-9625, Fax: 81-45-503-9586

COMMENT

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720;

Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda PLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (<http://pfweb.gsc.riken.go.jp/>) for further details.

FEATURES

source

1..1229

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/cultiVar="Columbia"

/db_xref="taxon:3702"

/chromosome="1"

/clone="RAF19-64-H16"

/notes="common name: thale cress"

1..1229

/gene="At1g02840/F22D16_30"

93..1004

/gene="At1g02840/F22D16_30"

/codon_start=1

/product="putative ribonucleoprotein SF-2"

/protein_id="BAC42991.1"

/db_xref="GI:26451794"

/translation="MSSRSRTVVGNLPGDIREREVEDLFSKVPVQIDLVPPRP PGTAFFEDDARDADAEALHGRDGYDEGHLRLVELAHGGRSDDDTNGSFGGGRGG RGRDGGSGRPSRRSEFVLVTGLPDSASWQDLKMRKGGDVCFSQVYRDARTGV VDTCTEDMKYALKLDLDFRNAFNGYVRYREYRSDRSRSPRSRSPRSRSPRSR GRVSRSRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRS SVQKEGSKSPSKPSAKSPHTHRSR"

BASE COUNT 318 a 256 c 328 g 327 t

ORIGIN

Alignment Scores:

Pred. No.: 4,26e-16 Length: 1229

Score: 143.00 Matches: 148

Percent Similarity: 68.40% Conservative: 10

Best Local Similarity: 64.07% Mismatches: 63

Query Match: 64.41% Indels: 10

DB: 8 Gaps: 1

US-10-014-927-19MOD_COPY_1_222 (1-222) x AK118379 (1-1229)

Qy 1 MetSerArgTyrPheAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20

Db 93 ATGAGCAGTCGTTCCAGTAGAACCGTGTACGTCGAAACCTTCCTCGCATATCCGTGAG 152

Qy 21 CysGluValGluAspLeuPheTyrLysTyrGlyProIleValAspIleAspLeuLysIle 40

Db 153 AGAGAGGTCGAAGATTGTTTTCAGTAAGTATGACCTGTCTTCAAAATGATTTGAAGTT 212

Qy 41 ProProArgProProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp 60

Db 213 CTTCCAAAGGCTCTGCTTATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 272

Qy 61 AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuValGluIleAla 80

Db 273 GCTATTTCATGCTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGCTG 332

Qy 81 HisGlyGlyArgArg----- 90

Db 333 CATGTTGGAGGCGTTTCATCAGATGATGCTCGGGGTAGTTTCAATGGTGGTGGTGGT 392

Qy 91 *****SerTyrSerAlaSerArgAlaProSerArgSerArgSerArgTyrArg 110

Db 393 GGTGGTGGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 452

Qy 111 ValLeuValThrGlyLeuProProSerAlaSerTyrGlnAspLeuIleAspHisMetArg 130

Db 453 GTTCTAGTCACAGGCTTGGCTTCTCATCTGCTTCTTGGCAAGATCTCAAGATCACATGCGT 512

Qy 131 LysAlaGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGlyVal 150

Db 513 AAAGGAGGCGATGCTCTGTTCTCGCAAGTGTACCGTGTGCTAGAGGGGCAACTGGAGTT 572

QY	151	ValAspTyrSerAsnTyrAspMetLysTyrAlaIleArgLysLeuAspAlaThrGlu	170
Db	573	GTGTGATTACACTGCTATGAGGACATGAAGTATGCGCTGAAGAAAGCTCGACGACACAGAG	632
QY	171	PheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgSerVal	190
Db	633	TTTCGAATAGCGTTTTCGAATGGATATGTCGGGTAGAGATATGATTCAAGGAAGGAT	692
QY	191	SerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgGlyProSer	210
Db	693	TCTAGGAGTCTAGCGGGAGAGATCTTATTCTAAGAGCCGACGCGCGGTGGACGA	752
QY	211	CysSerTyrSerSerLysSerArgSerValSer	221
Db	753	AGCGTGAGCGGAAGCAGGACGAGCAGCAGGAGC	785
RESULT 6			
AY085920			
LOCUS	AY085920	1232 bp	mRNA linear PLN 14-APR-2003
DEFINITION	Arabidopsis thaliana clone 19681 mRNA, complete sequence.		
ACCESSION	AY085920		
VERSION	AY085920.1	GI:21404630	
KEYWORDS	FLI CDNA.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Haas,B.J., Volfovsky,N., Town,C.D., Troughan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L. Full-length messenger RNA sequences greatly improve genome annotation		
AUTHORS	Genome Biol. 3 (6), RESEARCH0029 (2002)		
TITLE	Unpublished		
JOURNAL	22088475		
MEDLINE	12093376		
PUBMED	2 (bases 1 to 1232)		
REFERENCE	Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.		
AUTHORS	Full-length cDNA from Arabidopsis thaliana		
TITLE	Unpublished		
JOURNAL	3 (bases 1 to 1232)		
REFERENCE	Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.		
AUTHORS	Direct Submission		
TITLE	Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA		
JOURNAL			
COMMENT	This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.		
Location/Qualifiers			
1. .1232			
/organism="Arabidopsis thaliana"			
/mol_type="mRNA"			
/db_xref="taxon:3702"			
/clone="19681"			
94..1005			
/codon_start=1			
/product="SF2/ASF-like splicing modulator Srp30, putative"			
CDS			
FEATURES			
source			
protein_id="AAM63132.1"			
/db_xref="GI:21554051"			
/translation="MSSRSSTVYVGNLPDIREVEDELFSKYGPVVQIDLYKPPRRPGYAFVDFDADAEDAIHGRDGDYDFDGHRLRVELAHGRSSDDTRGSFNGGREGGGRGDGSGRSPSEFRVLTGLPSSASWQDLKHMRKGGDVCFQSVYRDKARTTGVVDYTCYEMKYALKLDDTEFNAPRNGYVRVREYDSRKDRSPSRGSRGSRGSRGSRGSRSPKAKSRRSPAKTSRSPGPRSKRSRSPSRSRSRSPPLPSVQKGSXSPKPSKPSAKSPIHTRSPSR"			
BASE COUNT 318 a 256 c 328 g 330 t			
ORIGIN			
Alignment Scores:			
Pred. No.: 4.28e-16 Length: 1232			
Score: 143.00 Matches: 148			
Percent Similarity: 68.40% Conservative: 10			
Best Local Similarity: 64.07% Mismatches: 63			
Query Match: 64.41% Indels: 10			
DB: 8 Gaps: 1			
US-10-014-927-19MOD_COPY_1_222 (1-222) x AY085920 (1-1232)			
QY	1	MetSerSerArgTTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys	20
Db	94	ATGAGCAGTCGTTTCGAGTAGAACCGTGTACGTGCGAAACCTTCTCGCGGATATCCGTGAG	153
QY	21	CysGluValGluAspLeuPheTyrLysTyrGlyProIleValAspIleAspLeuLysIle	40
Db	154	ACAGAGGTCGAAAGATTTGTTTCAGTAAGTATGACCTGTGTTTCAAAATTTGATTGAAGTT	213
QY	41	ProProArgProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp	60
Db	214	CCTCCAGGCCCTCCGTTTATGATTCGTTGAGTTTGTATGATCTCGGGATGCTGAAGAT	273
QY	61	AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla	80
Db	274	GCTATTTCATGGTCGTGATGCTATGACTTTGATGGCATCGTTTGAGGCGTGAATTGGCG	333
QY	81	HisGlyGlyArgArg-----	90
Db	334	CATGTTGGGAGCGCTTCATCAGATGATCTCGGGGTAGTTTCAATGGTGGTGGCGGTGT	393
QY	91	*****SerTyrSerAlaSerArgAlaProSerArgArgSerArgSerArgSer	110
Db	394	GGTGGTGGCGGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	453
QY	111	ValLeuValThrGlyLeuProProSerAlaSerTrpGlnAspLeuLysAspHisMetArg	130
Db	454	GTTCTAGTCACAGGCTTTCCTTCATCTGCTTTTGGCAAGATCTCAGGATCAGATCGGT	513
QY	131	LysAlaGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGlyVal	150
Db	514	AAAGGAGGCGATGCTGTTTCTCGCAAGTGTACCGGTGATGCTAGAGGGACAACACTGGAGTT	573
QY	151	ValAspTyrSerAsnTyrAspAspMetLysTyrAlaIleArgLysLeuAspAlaThrGlu	170
Db	574	GTTGATTACACCTGCTATGAGGACATGAAGTATGCGTGAAGAAAGCTCGACACAGAG	633
QY	171	PheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgSerVal	190
Db	634	TTTCGAATAGCGTTTTCGAATGGATATGTCGGGTAGAGATATGATTCAAGGAAGGAT	693
QY	191	SerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgGlyProSer	210
Db	694	TCTAGGAGTCTAGCGGGGAAGATCTTATTCTAAGAGCCGACGCGCGCGGTGGACGA	753
QY	211	CysSerTyrSerSerLysSerArgSerValSer	221
Db	754	AGCGTGAGCGGAAGCAGGACGAGCAGCAGGAGC	786
RESULT 7			
AY128338			
LOCUS	AY128338	1540 bp	mRNA linear PLN 07-AUG-2002
DEFINITION	Arabidopsis thaliana SF2/ASF-like splicing modulator Srp30,		

US-10-014-927-19MOD_COPY_1_222 (1-222) x AY128338 (1-1540)

QY 1 MetSerSerArgTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20
 DB 49 ATGAGCAGTCTGTCGAGTAGAACCGTGTACGTCGGAACCTTCTCGCGATATCCGTGAG 108
 QY 21 CysGluValGluAspLeuPheTyrLysTyrGlyProIleValAspIleAspLeuLysile 40
 DB 109 ACAGAGGTCGAAGATTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 168
 QY 41 ProProArgProProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp 60
 DB 169 CCTCAAGCCCTCTCTGTTATGATTCGTTGATTCGTTGATTCGTTGATTCGTTGATTCGTTGAT 228
 QY 61 AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla 80
 DB 229 GCTATTTCATGTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 288
 QY 81 HisGlyGlyArgArg-----***** 90
 DB 289 CATGTTGGGGCGGTCATCAGATGATCTCGGGGTAGTTTCAATGTTGGTGGCGGTGTT 348
 QY 91 *****SerTyrSerAlaSerArgAlaProSerArgArgSerAspTyrArg 110
 DB 349 GGTGTCGCGCGCGTGTGATGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGT 408
 QY 111 ValLeuValThrGlyLeuProSerAlaSerTrpGlnAspLeuLysAspHisMetArg 130
 DB 409 GTTCTAGTCACAGGCTTGCCTTCATCTGCTTCTTGGCAAGATCTCAAGGATCACATGCGT 468
 QY 131 LysAlaGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGlyVal 150
 DB 469 AAAGAGGCGATGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGT 528
 QY 151 ValAspTyrSerAsnTyrAspMetLysTyrAlaIleAlaGlyLysLeuAspAlaThrGlu 170
 DB 529 GTTGATTTACCTGTCATGAGGACATGAATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 588
 QY 171 PheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgSerVal 190
 DB 589 TTTCCGAATGGCTTTCCGATGATGTCGCGGTAGAGATATGATTCAGGAAGAT 648
 QY 191 SerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgSerArgSerVal 210
 DB 649 TCTAGGAGTCTAGCCGCGGGAAGATCTTATTTAAGACCGCAGCCGCGCGCGCGCGCGCGCG 708
 QY 211 CysSerTyrSerSerLysSerArgSerValSer 221
 DB 709 ACCGTGACCGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 741

RESULT 8
 ATHRIENUCP 1160 bp mRNA linear PLN 27-APR-1993
 LOCUS Arabidopsis thaliana ribonucleoprotein mRNA, complete cds.
 DEFINITION M98340
 ACCESSION M98340
 VERSION M98340.1 GI:166843
 KEYWORDS homologue; ribonucleoprotein; splicing factor; splicing factor
 SF-2.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE Lazar, G.
 AUTHORS Unpublished (1992)
 JOURNAL Original source text: Arabidopsis thaliana (strain Landsberg
 COMMENT erecta) (library: Lambda gtl1) root cDNA to mRNA.
 FEATURES
 source
 1..1160
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Landsberg erecta"

10266

putative (At1g02840) mRNA, complete cds.

AY128338
 VERSION AY128338.1 GI:22135917
 FLI CDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 1540)
 Tripp, M., Southwick, A., Karlin-Neumann, G., Nguyen, M., Miranda, M.,
 Palm, C.J., Bowser, L., Jones, T., Ban, J., Carninci, P., Chen, H.,
 Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
 Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.
 Direct Submission
 Submitted (01-JUL-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 e-mail for correspondence: arab@sequence.stanford.edu

The RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and
 sequenced the pUNI (ORF) clones using the RAFL cDNAs: Tripp, M.,
 Nguyen, M., Southwick, A., Karlin-Neumann, G., Lam, B., Miranda, M.,
 Palm, C.J., Bowser, L., Jones, T., Ban, J., Chen, H., Cheuk, R.,
 Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H.,
 Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.

Tripp, M. (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed equally
 to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W.
 (SSP/Stanford) contributed equally to this work as PIs.

Location/Qualifiers
 1..1540
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="RAFL04-19-E10 (R15578)"
 /notes="This clone is in pBluescript
 ecotype: Columbia"
 1..1540
 /gene="At1g02840"
 /note="synonym: F22D16.16"
 49..906
 /gene="At1g02840"
 /codon_start=1
 /product="SP2/ASP-like splicing modulator Srp30, putative"
 /protein_id="AA091541.1"
 /db_xref="GI:22135918"
 /translation="MSRSRTYVGNLPGDIRERVEDLFSKYGPWQIDLVPRP
 PGVAFEDDARDAEDAIHGDYDFDGHRLVELAHGRRSSDDTRGSFNGGRRGG
 RRGDGGSGRSEPRFLVTLGPSASWQDLKMRKGGDYCFQVYRDARTGTGV
 VDVTCDMKYALKLDDTEFRNAPNSGYVRYEYDNRKDSRSPSRGRYSRSR
 GRVSRSRSRSRSPSRSPKASRSRSPKSTSRSPGRSPSRSPSRPRMTVTETLHL
 DNNIISGL"

BASE COUNT 398 a 297 c 399 g 446 t

ORIGIN

Alignment Scores:
 Pred. No.: 7,01e-16 Length: 1540
 Score: 143.00 Matches: 148
 Percent Similarity: 68.40% Conservative: 10
 Best Local Similarity: 64.07% Mismatches: 63
 Query Match: 64.41% Indels: 10
 DB: Gaps: 1

FEATURES
 source

gene

CDS

CDS	/db_xref="taxon:3702"		Db	991 AGCGTGGCCGAGCAGGAGCAGGAGCAGGAGC 723	
	/tissue_type="root"				
	/tissue_lib="Lambda gt11"				
	31..942				
	/note="human splicing factor SF-2 homologue"			RESULT 9	
	/codon_start=1			LOCUS	
	/product="ribonucleoprotein"			AY056185	
	/protein_id="AAA32856.1"			Arabidopsis thaliana At4g02430 mRNA sequence.	
	/db_xref="GI:166844"			Arabidopsis thaliana (thale cress)	
	/translation="MSSRSSTRTVYVGNLPGDIRREVDELFSKYGVVVQIDLVKVPFRP			Arabidopsis thaliana	
PGYAFVFDADAEADAIHGRDGYDFDGRHLRVELAHGRRSSDDTRGSFNGGGRGGG		Arabidopsis thaliana			
RGGYAFVFDADAEADAIHGRDGYDFDGRHLRVELAHGRRSSDDTRGSFNGGGRGGG		Arabidopsis thaliana			
VNYTCYDMKVALKXLDTEPNAFNGYVREYDVRKDSRSPRSRGRSYKSRSRGR		Arabidopsis thaliana			
GRSVRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSPRSRSP		Arabidopsis thaliana			
SVQKGSKSP		Arabidopsis thaliana			
300 a 233 c 318 g 309 t		Arabidopsis thaliana			
BASE COUNT		Arabidopsis thaliana			
ORIGIN		Arabidopsis thaliana			
Alignment Scores:		Arabidopsis thaliana			
Pred. No.: 2.34e-15		Arabidopsis thaliana			
Score: 139.00		Arabidopsis thaliana			
Percent Similarity: 66.67%		Arabidopsis thaliana			
Best Local Similarity: 62.34%		Arabidopsis thaliana			
Query Match: 62.61%		Arabidopsis thaliana			
DB: 8		Arabidopsis thaliana			
US-10-014-927-19MOD_COPY_1_222 (1-222) x ATRHBNUPC (1-1160)		Arabidopsis thaliana			
QY	1 MetSerSerArgTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20	Arabidopsis thaliana			
Db	31 ATGAGCAGCTGTTGCGATGAGACCGGTGATGCGGAACTTCTCGGCGATATCCGTGAG 90	Arabidopsis thaliana			
QY	21 CysGluValGluAspLeuPheTyrLysTyrGlyProIleValAspIleAspLeuLysIle 40	Arabidopsis thaliana			
Db	91 AGAGAGCTCGAAGATTGTTTCAGTAGTAGGACCTGTTGTTCAATTGATTGAGGTT 150	Arabidopsis thaliana			
QY	41 ProProArgProProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp 60	Arabidopsis thaliana			
Db	151 CTTCCAGGCTCTCTGTTATGATTCATTCGTTGATGATGATGATGATGATGATGATGAT 210	Arabidopsis thaliana			
QY	61 AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuValGluIleAla 80	Arabidopsis thaliana			
Db	211 GCTATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 270	Arabidopsis thaliana			
QY	81 HisGlyGlyArgArg-----***** 90	Arabidopsis thaliana			
Db	271 CATGTTGGAGGCGTTTCATCAGTAGTAGTACCGGGTAGTTTCAATGTTGTTGCGCGTGGT 330	Arabidopsis thaliana			
QY	91 *****SerTyrSerAlaSerArgAlaProSerArgArgSerArgTyrArg 110	Arabidopsis thaliana			
Db	331 GGTGGTGGCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 390	Arabidopsis thaliana			
QY	111 ValLeuValThrGlyLeuProProSerAlaSerTyrGlnAspLeuLysAspHisMetArg 130	Arabidopsis thaliana			
Db	391 GTTCTAGTCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 450	Arabidopsis thaliana			
QY	131 LysAlaGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGlyVal 150	Arabidopsis thaliana			
Db	451 AAAGAGGCGCATGCTGTTTCTCGCAAGTGTACCGTGTAGTAGGAGGACCACTGGAGTT 510	Arabidopsis thaliana			
QY	151 ValAspTyrSerAsnTyrAspMetLysTyrAlaIleArgLysLeuAspAlaThrGlu 170	Arabidopsis thaliana			
Db	511 GTTGATTACACCTGCTATGAGGACATGAGTATGCGTGAAGAGTCTGACGACACAGAG 570	Arabidopsis thaliana			
QY	171 PheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgSerVal 190	Arabidopsis thaliana			
Db	571 TTTTCGAAATCGCTTTTCGAAATCGATATGTCGCGGTTAGAGATATGATTCAGAGGAGGAT 630	Arabidopsis thaliana			
QY	191 SerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgGlyProSer 210	Arabidopsis thaliana			
Db	631 TCTAGGAGTCTCTAGCGGGAAGATCTTATTCTAGACCGGACCGCGCGTGGACGA 690	Arabidopsis thaliana			
QY	211 CysSerTyrSerSerLysSerArgSerValSer 221	Arabidopsis thaliana			

FEATURES	source	/organism="Arabidopsis thaliana"	
		/mol_type="mRNA"	
		/db_xref="taxon:3702"	
		/chromosome="4"	
		/clone="RAF108-15-J15 (R11393)"	
		/note="This clone is in a modified pBluescript vector"	
		Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.	
		Location/Qualifiers	
		1..1523	
		The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAF1 cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Toriumi, M., Wallender, E.K., Wong, C., Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Shinn, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W., Ecker, J.R. and Theologis, A.	
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.			
Location/Qualifiers			
1..1523			
/organism="Arabidopsis thaliana"			
/mol_type="mRNA"			
/db_xref="taxon:3702"			
/chromosome="4"			
/clone="RAF108-15-J15 (R11393)"			
/note="This clone is in a modified pBluescript vector"			

FEATURES

source

1. 1523

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="4"

/clones="RAFL08-15-J15 (R11393)"

/note="This clone is in a modified pBluescript vector"

DEFINITION		Arabidopsis thaliana At4g02430 mRNA for unknown protein, complete cds, clone: RAPL19-32-J05.	
ACCESSION		AKI18074	
VERSION		AKI18074.1 GI:26451202	
KEYWORDS		FLI CDNA; CAP trapper.	
SOURCE		Arabidopsis thaliana (thale cress)	
ORGANISM		Arabidopsis thaliana	
REFERENCE		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi	
AUTHORS		Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.	
TITLE		Arabidopsis thaliana full-length cDNA	
JOURNAL		Published Only in Database (2002)	
REFERENCE		2 (bases 1 to 2059)	
AUTHORS		Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.	
TITLE		Direct Submission	
JOURNAL		Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: msek@gsc.riken.go.jp)	
COMMENT		URL: http://pfweb.gsc.riken.go.jp, tel:81-45-503-9625, fax:81-45-503-9586	
FEATURES		An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://pfweb.gsc.riken.go.jp/) for further details.	
source		Location/Qualifiers	
gene		1..2059	
CDS		/organism="Arabidopsis thaliana"	
		/mol_type="mRNA"	
		/cultivar="Columbia"	
		/db_xref="taxon:3702"	
		/chromosome="4"	
		/clone="RAPL19-32-J05"	
		/note="common name: thale cress"	
		1..2059	
		/gene="At4g02430/T14P8_21"	
		153..689	
		/gene="At4g02430/T14P8_21"	
		/codon_start=1	
		/product="unknown protein"	
		/protein_id="BAC42705.1"	
		/db_xref="GI:26451203"	
		/translation="MSRSRTIYVGNLPGDIREREVEDLFSKYGVVVIDLKTIPP	
		PGYAFVFEADADDAYGRDGYDFDGHLLVELAHGRRSSHDARGSGRGRGR	
		GGDGGRRGRSPRSRYVVGSLPSSASQDLKDHMRKGVEYFCSQVFRDGRGTG	
		IVDVTSDMKYAIKKAR"	
BASE COUNT		558 a 365 c 506 g 630 t	
ORIGIN			
Alignment Scores:		2,56e-13 Length: 2059	
Pred. No.:		131.50 Matches: 150	
Score:		68.67% Conservative: 10	
Percent Similarity:		64.38% Mismatches: 61	
Best Local Similarity:		59.23% Indels: 12	
Query Match:		8 Gaps: 1	
DB:			
US-10-014-927-19MOD_COPY_1_222 (1-222) x AY056185 (1-1523)			
Qy	1	MetSerSerArgTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20	
Db	49	ATGAGCAGCGTTCGAGTAGAAGATTACCTCGGAACCTCCCGCGATATCCGTGAA 108	
Qy	21	CysGluValGluAspLeuPheTyrLysTyrGlyProIleValAspIleAspLeuTysIle 40	
Db	109	AGAGAACTTCAAGACTTGTTCAGTAGATGGACCTGTTCCTCAATCGATTGAAGATT 168	
Qy	41	ProProArgProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAsp 60	
Db	169	CCGCCAGGCGCTCCAGGCTATGCTGCTGAGTTTGAGGATCTCTGATGCTGATGAT 228	
Qy	61	AlaIleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla 80	
Db	229	GCAATTTATGCCGTGATGTTATGACTTTTGATGGGCATCATTTACGGGTGGAACTAGCT 288	
Qy	81	HisGlyGlyArg----- 89	
Db	289	CATGGTGGAGGGCTTCATCAGATGATGCACGGCGGTAGTTATAGTGTCTGGTCTGGGC 348	
Qy	90	*****SerTyrSerAlaSerArgAlaProSerArgArgSerAspTyr 109	
Db	349	GCTCGTGTGTTGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 408	
Qy	110	ArgValLeuValThrGlyLeuProProSerAlaSerTyrGlnAspLeuLysAspHisMet 129	
Db	409	CGCGTTGTAGTGTGAGTTTGCCTTCATCTCGCTCCCTGGCAAGACCTCAAGCATCATG 468	
Qy	130	ArgIleValGlyAspValCysPheSerGluValPheProAspArgLysGlyMetSerGly 149	
Db	469	CGTAAAGAGGAGGAAGTTTGTCTTCTCAAGTGTTCGTGATGGTAGAGTACAACTGGA 528	
Qy	150	ValValAspTyrSerAspMetLysTyrAlaIle-ArgLysLeuAspAlaThr 169	
Db	529	ATTGTAGATTATACCAGTACAGGACATGAATATGCGATAAAGCTCGATCAGAC 588	
Qy	169	rGluPheArgAsnAlaPheSerSerAlaTyrIleArgValArgLysTyrGluSerArgSer 189	
Db	589	AGAGTTTCGAATGCGTTTCTCATGAATATGTTCCGGTTAGAGATATGATTCAAGGAG 648	
Qy	189	rValSerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgGlyPr 209	
Db	649	GGATTCCAGGAGGCCCGCCGCGGAGATCTTATTTAAGAGCCGCGCGGTGGACGGAG 708	
Qy	209	oSerCysSerTyrSerSerLysSerArgSerValSer 221	
Db	709	TCCAGCGGTAGTGTAGTCCGCGCAGACAGACGAGC 745	
RESULT 10			
AKI18074			
LOCUS		AKI18074 2059 bp mRNA linear PLN 06-DEC-2002	
Qy		1 MetSerSerArgTrpAsnArgThrIleTyrValGlyAsnLeuProGlyAspIleArgLys 20	
Db		153 ATGAGCAGCGTTCGAGTAGAAGATTACCTCGGAACCTCCCGCGATATCCGTGAA 212	

Db 213 AGAAGAGTTGAGACTTGTTCAGTAAAGTATGACCTGTGTTCAAAATCGATTTGAAGATT 272
QY 41 PProProArgProProGlyTyrAlaPheValGluPheGluAspProArgAspAlaAspAsp 60
Db 273 CGCGCGAGGCGCTCCAGGCTATGATTCGTGAGTTTGAAGATGCTGCTGATGCTGATGAT 332
QY 61 AlalleTyrGlyArgAspGlyTyrAspPheAspGlyCysArgLeuArgValGluIleAla 80
Db 333 GCAATTTATGCGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 392
QY 81 HisGlyGlyArg-----***** 89
Db 393 CATGTGGGAGCGTTTCATCATCATGATGACCGGTAGTATTAGTGGTGGTGGTGGG 452
QY 90 *****SerTyrSerAlaSerArgAlaPProSerArgArgSerArgSerArgTyr 109
Db 453 GGTCTGT 512
QY 110 ArgValLeuValThrGlyLeuProProSerAlaSerTyrGlnAspLeuLysAspHisMet 129
Db 513 CGCGTTGTAGTGTGAGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
QY 130 ArgLysAlaGlyAspValCysPheSerGluValPheProAspArgGlyGlyMetSerGly 149
Db 573 CGTAAGGAGGAGAAAGTTTCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 632
QY 150 ValValAspTyrSerAsnTyrAspMetLysTyrAlaIle-ArgLysLeuAspAlaLth 169
Db 633 ATTGTAGATTATACAGCTACAGAGCATGAATATGCGATAAAGCTCGATGACAC 692
QY 169 rGluPheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyrGluSerArgse 189
Db 693 AGAGTTTCGGAATGCGTTTCTCATGATATGTTCTCGGTAGAGATATGATTCAGGAG 752
QY 189 rValSerArgSerProAspAspSerLysSerTyrArgSerArgSerArgSerArgGlyPr 209
Db 753 GGATTCGAGAGCGCGCGTGGAGATCTCTATTCTAAGAGCGCGAGCGCGCGAGCGGAG 812
QY 209 oSerCysSerTyrSerLysSerArgSerValSer 221
Db 813 TCCAGCGCGTGTCTGTAGCGGAGCAGAGCAGGAGC 849

BD094070 1428 bp DNA linear PAT 27-AUG-2002
LOCUS Shear stress-responsive DNAs.
DEFINITION BD094070
ACCESSION BD094070.1 GI:22639658
VERSION WO 0125427-A/31.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1428)
AUTHORS Nojima,H., Yoshisue,H., Obayashi,M., Ota,T., Kawabata,A.,
Sakurada,K., Kuga,T., Sekine,S., Nakamura,Y. and Sugan,S.
TITLE Shear stress-responsive DNAs
JOURNAL KYOWA HAKKO KOGYO CO LTD,HIROSHI NOJIMA,HAJIME YOSHISUE, MASAYA
OBAYASHI, TOSHIO OTA,AYAKO KAWABATA,KAZUHIRO SAKURADA,TETSURO KUGA,
SUMIO SEKINE, YUSUKE NAKAMURA, SUMIO SUGANO
COMMENT OS Homo sapiens (human)
PN WO 0125427-A/31
PD 12-APR-2001
PF 02-OCT-2000 WO 2000JP006840
PR 01-OCT-1999 JP 99P 280976
PI HIROSHI NOJIMA,HAJIME YOSHISUE,MASAYA OBAYASHI,TOSHIO OTA, PI
AYAKO KAWABATA,
PI KAZUHIRO SAKURADA,TETSURO KUGA,SUSUMU SEKINE,YUSUKE NAKAMURA,
SUMIO SUGANO
PC C12N15/12,C07K14/435,C07K16/18,C12P21/02,C12Q1/68,A61K38/00,
PC A61K39/395,

PC A61K48/00,A61P9/10,G01N33/50,G01N33/53
CC
FH Key Location/Qualifiers
FT CDS (125)..(868).

FEATURES
source
1..1428
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606" 391 t
BASE COUNT 341 a 296 c 400 g
ORIGIN

Alignment Scores: 2.04e-06 Length: 1428
Pred. No.: 95.00 Matches: 108
Score: 56.46% Conservative: 10
Best Local Similarity: 51.67% Mismatches: 85
Query Match: 42.79% Indels: 6
DB: Gaps: 2

US-10-014-927-19MOD_COPY_1_222 (1-222) x BD094070 (1-1428)

QY 9 IleTyrValGlyAsnLeuProGlyAspIleArgLysCysGluValGluAspLeuPheTyr 28
Db 176 ATCTAGTGGTAACTTACCTCCAGACATCCGACATCCGAAACCAAGACATTGAGACGTGTTCTAC 235
QY 29 LysTyrGlyProIleValAspIleAspLeuLysIleProProArgProGlyTyrAla 48
Db 236 AATACGGCGCTATCCGACATCCGACCTCAAGATCCGCGGGGAGCCGCCCTTCGCC 295
QY 49 PheValGluPheGluAspProArgAspAlaAspAspAlaIleTyrGlyArgAspGlyTyr 68
Db 296 TTCGTTGAGTTCGAGGACCCGCGAGACGCGGAGACGCGGTGTATGTCGCGACGCGCTAT 355
QY 69 AspPheAspGlyCysArgLeuArgValGluIleAlaHisGlyGlyArg-----Arg 85
Db 356 GATTACGATGGGTACCGTCTCGCGGTGGAGTTTCCTCGAAGCGCGCGTGGAAACAGCCCGA 415
QY 86 *****SerTyrSerAlaSerArgAlaProSerArg 105
Db 416 GCGCGCGCGGGGTGGAGTGGCGGAGTCCCGAGGTCCGATGCGGCCCATCCAGG 475
QY 106 ArgSerAspTyrArgValLeuValThrGlyLeuProProSerAlaSerTrpGlnAspLeu 125
Db 476 CGGTCTGAAAACAGAGTGTGTTCTCTGACTGCTCCCAAGTGGAAAGTTGGCAGGATTTA 535
QY 126 LysAspHisMetArgLysAlaGlyAspValCysPheSerGluValPheProAspArgLys 145
Db 536 AAGGATCATCGTCAAGCAGGTGATGTATGTTATGCTGTATGCTGTATGCTGTATGCTGTAT 589
QY 146 GlyMetSerGlyValValAspTyrSerAsnTyrAspAspMetLysTyrAlaIleArgLys 165
Db 590 ---GGCAGCTGGTGTCTGTGGAGTTGTACGGAAAGAGATATGACCTATGCACTTCGAAA 646
QY 166 LeuAspAlaThrGluPheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyr 185
Db 647 CTGGATAACACTAAGTTTAGATCTCATGAGGGGAAACTGCCTACATCCGGGTAAAGTT 706
QY 186 GluSerArgSerValSerArgSerProAspAspSerLysSerTyrArgSerArgSerArg 205
Db 707 GATGGGCCCAAGAGTCCAAAGTTATGAGATCTCGATCTCGAAGCGGTAGTCTGTAGCAGA 766
QY 206 SerArgGlyProSerCysSerTyrSer 214
Db 767 AGCCGTAGCAGAGCAACACGAGGAGT 793

RESULT 12
HUMSF2P33 1428 bp mRNA linear PRI 24-NOV-1993
LOCUS Human SF2p33 mRNA, complete cds.
DEFINITION M69040
ACCESSION M69040.1 GI:338046
VERSION M69040.1
KEYWORDS SF2p33.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Krainer, A.R., Mayeda, A., Kozak, D., and Binns, G.
TITLE Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, UI 70K, and Drosophila splicing regulators
JOURNAL Cell 66 (2), 383-394 (1991)
MEDLINE 91309150
PUBMED 1830244
COMMENT Original source text: Homo sapiens cDNA to mRNA.
Comments: SF2p33 is an essential pre-mRNA splicing factor, which can also affect alternative 5' splice site selection in vitro by stimulating the use of proximal 5' splice sites.
(Krainer, A.R., Conway, G.C., and Kozak, D. (1990) Purification and Characterization of SF2, a Human Pre-mRNA Splicing Factor. Genes Dev. 4, 1158-1171; Krainer, A.R., Conway, G.C., and Kozak, D. (1990) The Essential Pre-mRNA Splicing Factor SF2 Influences 5' Splice Site Selection by Activating Proximal Sites. Cell 62, 35-42). This factor is also known as ASF (Ge, H., and Manley, J.L. (1990) A protein factor, ASF, Controls Alternative Splicing of SV40 Early Pre-mRNA In Vitro. Cell 62, 25-34). SF2 p33 consists of two polypeptides of apparent molecular weight 33 kd, approximately, although their predicted molecular weight is 27,744 daltons. The two forms appear to differ by the extent of post-translational modification, which includes phosphorylation. SF2p33 binds RNA and promotes the annealing of complementary RNAs. It is required for assembly of pre-spliceosome complexes. The N-termini of the HeLa polypeptides are blocked.
Location/Qualifiers
1. 1428
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="female"
125..871
/function="pre-mRNA splicing factor"
/codon_start=1
/product="SF2p33"
/protein_id="AAA03476.1"
/db_xref="GI:338047"
/translation="MSGGGVYVGNLPPDIRTKDIEDVYKYGAIRD IDLKNRGGPPFAVFEDPRDAEDAVYGDYDGYDYLRFVFPSSRGRTGGGGG GGGGAGRGVGPSSRSRSENVVYSGLPFGSGWQDLKDHREAGDVYADYRDGVV EFVKEDMTAVAKLQNTKPSRHEGTAYIRVKGDPSPSYGSRSRSRSRSRSR SRSRSRSPSRSGSPRYSPSRSRST"
872..1428
3'UTR 341 a 296 c 400 g 391 t
ORIGIN
Alignment Scores:
Pred. No.: 2,04e-06 Length: 1428
Score: 95.00 Matches: 108
Percent Similarity: 56.46% Conservative: 10
Best Local Similarity: 51.67% Mismatches: 85
Query Match: 42.79% Indels: 6
Gaps: 2
US-10-014-927-19MOD_COPY_1_222 (1-222) x HUMSF2p33 (1-1428)
Qy 9 IleTyrValGlyAenLeuProGlyAspIleArgLysCysGluValGluAspLeuPheTyr 28
Db 176 ATCTACGTGGGTAACTTACCTCCAGACATCCGAACCAAGACATTGAGGACGTGTTCTAC 235
Qy 29 LysTyrGlyProIleValAspIleAspLeuLysIleProArgProProGlyTyrAla 48
Db 236 AAATACGGCGCTATCCCGACATCGACTCAAGATCGCGCGGGAGCGCCCTTCGCC 295
Qy 49 PheValGluPheGluAspProArgAspAlaAspAlaIleTyrGlyArgAspGlyTyr 68
Db 296 TTCGTTAGTTCGAGGACCGCGAGACCGCGGTGTATGTCGCGACGGCTAT 355

Qy 69 AspPheAspGlyCysArgLeuArgValGluIleAlaHisGlyGlyArg-----Arg 85
Db 356 GATTACGATGGTACCGGTCCTCGGGTGAGTTCTCGAAGCGCGCGTGGAAACAGCGCA 415
Qy 86 *****SerTyrSerAlaSerArgAlaProSerArg 105
Db 416 GCGCGCGCGGGGTGGAGTGGCGAGCTCCGAGGTGCTATGCGCCCTCCAGG 475
Qy 106 ArgSerAspTyrArgValLeuValThrGlyLeuProProSerAlaSerTyrGluAspLeu 125
Db 476 CGGTCGTAACACAGAGTGGTGTCTCTGGACTGCTCCAGTGGGAAGTGGCAGATT 535
Qy 126 LysAspHisMetArgLysAlaGlyAspValCysPheSerGluValPheProAspArgLys 145
Db 536 AAGGATCATACCGGTGAGCGAGGTGATGTATGCTGATGTTTACCGAGAT----- 589
Qy 146 GlyMetSerGlyValValAspTyrSerAsnTyrAspAspMetLysTyrAlaIleArgLys 165
Db 590 ---GGCATCTGGTGTGGAGTTGTACGGAAGAGATATGACTATGCACTTCGAGTT 646
Qy 166 LeuAspAlaThrGluPheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyr 185
Db 647 CTGGATACATAGTTTAGATCTCATGAGGAGAACTGCTACATCCGGGTAAAGTT 706
Qy 186 GluSerArgSerValSerArgSerProAspAspSerLysSerTyrArgSerArgSerArg 205
Db 707 GATGGCGCCAGAGTCCCAAGTTATGGAAGATCTCGATCTCGAAGCGGTAGTCGTAGC 766
Qy 206 SerArgGlyProSerCysSerTyrSer 214
Db 767 AGCCGTAGCAGAACACAGCAGGAGT 793

RESULT 13
LOCUS BC010264
DEFINITION Homo sapiens, splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor), clone MGC:5228 IMAGE:2900101, mRNA, complete cds.
ACCESSION BC010264
VERSION BC010264.1 GI:16307433
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2708)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 3 Row: e Column: 5
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5902075.

102(b) 10

Db	618	GATGGCCCCAGCAAGTCCAAAGTTATCGAAGATCTCGAAGCGGTAGTCGTAGCACA	677
Qy	206	SerArgGlyProSerCysSerTyrSer	214
Db	678	AGCCGTAGCAGAACCAACAGCAGGAGT	704
RESULT 14			
AB062124			
LOCUS			
DEFINITION			
AB062124			
ACCESSION			
VERSION			
AB062124.1		GI:21104371	
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE			
1	Shichijo, S. and Itoh, K.		
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
2	(bases 1 to 2765)		
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1. 2765			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/cell_lines="SW620"			
/cell_type="colon adenocarcinoma"			
1. .2765			
/gene="OK/SW-cl.3"			
91..837			
/note="similar to splicing factor SF2p33"			
/codon_start=1			
/protein_id="BAB93456.1"			
/db_xref="GI:21104372"			
/translation="MSGGGVIRGPAAGNDRIYVGNLPDITKDIEDVFYKYGAIRD			
IDLKVRGPPFAFVFEEDPRDAEDAVYGRDGYDGLRVRFEFRSGRGTRGGGGG			
GGGAPRGYGPSSNRVSVGLPSSGQDLKDHREAGDVCVADYVDGTGVV			
GGGAPRGYGPSSNRVSVGLPSSGQDLKDHREAGDVCVADYVDGTGVV			
EFVRKEDMTYAVRKLDNTKFRSHEGETAYIRVKDGPSPSYGRSRSRSTRSR			
SNRSRSYSPRSRSGSPYSPHRSRST"			
BASE COUNT	759	a 500	c 652
ORIGIN			
Alignment Scores:			
Pred. No.:	8.77e-06	Length:	2765
Score:	95.00	Matches:	108
Percent Similarity:	56.46%	Conservative:	10
Best Local Similarity:	51.67%	Mismatches:	85
Query Match:	42.79%	Indels:	6
DB:	9	Gaps:	2
US-10-014-927-19MOD_COPY_1_222 (1-222) x AB062124 (1-2765)			
Qy	9	IleTyrValGlyAsnLeuProGlyAspLeuArgLysCysGluValGluAspLeuPheTyr	28
Db	142	ATCTAGTGGTAACTTACTCTCAGACATCCGACCAAGACATTCGAGCGTGTCTTAC	201
Qy	29	LysTyrGlyProIleValAspLeuLysIleProProArgProGlyTyrAla	48
Db	202	AATACGGCGCTATCCGCGACATCGACCTCAAGATCGCGCGGGGACCGCCCTTCGCC	261
Qy	49	PheValGluPheGluAspProArgAspAlaAspAspAlaIleTyrGlyArgAspGlyTyr	68
Db	262	TTCGTGAGTTCAGAGACCGCGAGACCGCGAGACCGCGAGACCGGTGTATGTCGCGAGCGGTAT	321

QY 69 AspPheaspGlyCysArgLeuArgValGluLeuAlaHisGlyGlyArg-----Arg 85
Db 322 GATTACCATGGTACCCTCTCGGGTGGAGTTCTCGAAGCGCGCGTGGACAGCGCGA 381
QY 86 *****SerTyrSerAlaSerArgAlaProSerArg 105
Db 382 GCGCGCGCGGGGTGAGGTGGCGAGTCCCGAGTGGCTATGCGCCCGCATCGAGG 441
QY 106 ArgSerAspTyrArgValLeuValThrGlyLeuProProSerAlaSerTrpGlnAspLeu 125
Db 442 CGGTCTGAAACACAGAGTGGTGTCTCTGGAGTCTCCAGTGGAGTGGCAGGATTTA 501
QY 126 LysAspHisMetArgLysAlaGlyAspValCysPheSerGluValPheProAspArgLys 145
Db 502 AAGGATCACATGCTGAGCAGGTCATCTATGCTGATGTTTACCGAGAT----- 555
QY 146 GlyMetSerGlyValValAspTyrSerAsnTyrAspAspMetLysTyrAlaLeuArgLys 165
Db 556 ---GGCACTGGTGGTGGAGTTGTACGGAAGAGATATGACCTATGAGTTCGAAAA 612
QY 166 LeuAspAlaThrGluPheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyr 185
Db 613 CTGTATACACTAAGTTTAGTCTCATGAGGAGAACTGCTACATCCGGTTAAAGTT 672
QY 186 GluSerArgSerValSerArgSerProAspAspSerLysSerTyrArgSerArgSerArg 205
Db 673 GATGGCCCAAGTCCCAAGTTATGGAAGATCTCGATCTCGAAGCGGTAGTGTAGCAGA 732
QY 206 SerArgGlyProSerCysSerTyrSer 214
Db 732 AGCCGTAGCAGAGCAACAGCAGGAGT 759

RESULT 15

BC046773 2878 bp mRNA linear ROD 14-FEB-2003
LOCUS
DEFINITION Mus musculus, Similar to splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor), clone MGC:61417 IMAGE:5708288, mRNA, complete cds.

ACCESSION BC046773.1 GI:28386235
VERSION
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2878)
Strausberg, R.
Direct Submission
Submitted (13-FEB-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: University of Iowa, Dr. M. Bento Soares and Dr. Thomas L. Casavant.
Web site: <http://genome.uiowa.edu>
Contact: bento-soares@uiowa.edu; tom-casavant@uiowa.edu
Bonaldi, M.F., Akabogu, I., Bair, T., Bair, J., Crouch, K., Davis, A., Finkler, K., Keppel, C., Kucaba, T., Lebeck, M., Melo, A., Schaefer, K., Scheetz, T., Smith, C., Snir, E., Tack, D., Trout, K., Walters, J., Casavant, T., Soares, M.B.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: Plate: Row: Column: 0
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF

analysis, GenomeScan gene prediction.

FEATURES
source
1..2878
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="MGC:61417 IMAGE:5708288"
/tissue_type="Brain, mouse 15.5 dpc"
/clone_lib="NIH BMAP_EW0"
/lab_host="DH10B"
/note="Vector: pYX"
99..845
/codon_start=1
/product="Similar to splicing factor, arginine/serine-rich 1 (splicing factor 2, alternate splicing factor)"
/protein_id="AAH46773.1"
/translation="MSGGCVTRGPGAGNDCRIYVGNLPDITKIDIEOVFKYGAIRD IDLKNRRGGPPFAFVEFDPDASDAVYGRDGYDYGRLRVEFPRSGRGTGGGGG GGGAGRGYGPFRSRNRYVSGLPSPSGQDLKDMRAGADVCYADVVDGTVV EFVRKEDMTYAVRKLDTNFKRSHGSETAYIRKVDGPRSPSYGRSRSRSRSR SRSNRSRSPSRSPSRSPSRSPSRST"
BASE COUNT 775 a 530 c 582 g 891 t
ORIGIN

CDS

Alignment Scores:
Pred. No.: 9,588-06 Length: 2878
Score: 95.00 Matches: 108
Percent Similarity: 56.46% Conservative: 10
Best Local Similarity: 51.67% Mismatches: 85
Query Match: 42.79% Indels: 6
DB: 10 Gaps: 2

US-10-014-927-19MOD_COPY_1_222 (1-222) x BC046773 (1-2878)

QY 9 IleTyrValGlyAsnLeuProGlyAspLeuArgLysCysGluValGluAspLeuPheTyr 28
Db 150 ATCTACGTGGGTAACTACCTCCGATATCCGACCAAGGACATCGAGACGCTGTTTAC 209
QY 29 LysTyrGlyProIleValAspLeuAspLeuLysIleProProArgProProGlyTyrAla 48
Db 210 AAATACGCGCCATCCGACATCGACCTGAAGAACCCCGCGGGGACCGCCCTTCGCC 269
QY 49 PheValGluPheGluAspProArgAspAlaAspAspAlaIleTyrGlyArgAspGlyTyr 68
Db 270 TTCGTGTAGTTCGAGACCCCGAGACCGGAGATCGGTGTACGTCGCGACGCTAC 329
QY 69 AspPheAspGlyCysArgLeuArgValGluIleAlaHisGlyGlyArg-----Arg 85
Db 330 GACTACGAGGTACCGCTCGGTAGATTTCCCGAGCGCGCGGGACCGCGCGA 389
QY 86 *****SerTyrSerAlaSerArgAlaProSerArg 105
Db 390 GCGCGCGCGGGGTGGAGCGCGCGCGCGAGAGCGCGCTATGCGCGCGCTCCAGG 449
QY 106 ArgSerAspTyrArgValLeuValThrGlyLeuProProSerAlaSerTrpGlnAspLeu 125
Db 450 CGGTCTGAAACACAGAGTGGTGTCTCTGGAGTCTCCAGTGGAGTGGCAGGACTTA 509
QY 126 LysAspHisMetArgLysAlaGlyAspValCysPheSerGluValPheProAspArgLys 145
Db 510 AAGGATCACATGCTGAGCAGGTCATCTATGTTACGCTGATGTTTACCGAGAT----- 563
QY 146 GlyMetSerGlyValValAspTyrSerAsnTyrAspAspMetLysTyrAlaLeuArgLys 165
Db 564 ---GGCACTGGTGGTGGAGTTGTACGGAAGAGATATGACGTATGACGTTCGAA 620
QY 166 LeuAspAlaThrGluPheArgAsnAlaPheSerSerAlaTyrIleArgValArgGluTyr 185
Db 621 CTGTATACACTAAGTTTAGTCTCATGAGGAGAACTGCTACATCCGGTTAAAGTT 680
QY 186 GluSerArgSerValSerArgSerProAspAspSerLysSerTyrArgSerArgSerArg 205

Db 681 GATGGCCAGAGTCCAAAGTTATGGAAGATCTCGATCTCGAAGCCGTAGTCGTAGCAGA 740
QY 206 SerArgGlyProSerCysSerTyrSer 214
Db 741 ACCCGTAGCAGAGCAACACAGCAGGAGT 767

Search completed: February 4, 2004, 23:05:47
Job time : 3518 secs